U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

AD-A027 885

SOME ASYMPTOTIC RESULTS FOR OCCUPANCY PROBLEMS

WISCONSIN UNIVERSITY

PREPARED FOR
ARMY RESEARCH OFFICE

APRIL 1976

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 1600	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) SOME ASYMPTOTIC RESULTS FOR OCCUPANCY PROBLEMS		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Lars Holst		8. CONTRACT OR GRANT NUMBER(s) DAAG29-75-C-0024
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709		12. REPORT DATE April 1976
		13. NUMBER OF PAGES ~~14~~ 17
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Occupancy problems, coupon collectors problem, limit theorems

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Consider a situation in which balls are falling into N cells with arbitrary probabilities. Limit distributions for the number of empty cells are considered when $N \to \infty$ and the number of balls $n \to \infty$ so that $n/N \to \infty$. Limit distributions for the number of balls to achieve exactly b empty cells are obtained when $N \to \infty$ for b fixed or $b \to \infty$ so that $b/N \to 0$.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

MRC Technical Summary Report #1600

SOME ASYMPTOTIC RESULTS FOR OCCUPANCY PROBLEMS

Lars Holst

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

April, 1976

Received September 19, 1975

Sponsored by

U. S. Army Research Office
P.O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

SOME ASYMPTOTIC RESULTS FOR OCCUPANCY PROBLEMS

Lars Holst

Technical Summary Report # 1600
April 1976

ABSTRACT

Consider a situation in which balls are falling into N cells with arbitrary probabilities. Limit distributions for the number of empty cells are considered when $N \to \infty$ and the number of balls $n \to \infty$ so that $n/N \to \infty$. Limit distributions for the number of balls to achieve exactly b empty cells are obtained when $N \to \infty$ for b fixed or $b \to \infty$ so that $b/N \to 0$.

AMS(MOS) Classification: Primary 60F05, Secondary 60C05

Key Words: Occupancy problems; coupon collectors problem; limit theorems.

Work Unit No. 4 (Probability, Statistics and Combinatorics)

A

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024

SOME ASYMPTOTIC RESULTS FOR OCCUPANCY PROBLEMS

Lars Holst

1. Introduction.

Suppose that balls are thrown independently of each other into N cells, so that each ball has the probability p_k of falling into the kth cell, $p_1 + \ldots + p_N = 1$. Let Y_n denote the number of empty cells after n throws and let T_b denote the throw on which for the first time exactly b cells remain empty, $0 \leq b < N$. The symmetrical case $p_1 = \ldots = p_N = 1/N$ is discussed in e.g. Feller (1968), see occupancy or waiting time problems.

Depending on how $b, n, N \to \infty$, different asymptotic distributions for Y_n and T_b can be obtained, see e.g. Holst (1971) and for the symmetric case see e.g. Samuel-Cahn (1974). In this paper some remaining problems are investigated for the nonsymmetrical case.

To give precise meanings of the limits obtained, double sequences e.g. $(p_{kN})_N, (Y_{nN})_N$ are considered. But in order to simplify the notation the extra index N will usually be omitted.

2. A bounded number of empty cells.

The following limit theorem for Y_n, the number of empty cells after n throws, was proved by Sevastyanov (1972).

Theorem 1. If the p's are such that

$$(2.1) \qquad \max_{1 \leq k \leq N} (1 - p_k)^n \to 0$$

and

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

$$E(Y_n) = \sum_{k=1}^{N} (1 - p_k)^n \to m < \infty , \tag{2.2}$$

then

$$P(Y_n = y) \to m^y \cdot e^{-m}/y! \ , \tag{2.3}$$

or equivalently

$$Y_n \Rightarrow Po(m) \ , \quad \text{when} \quad N \to \infty . \tag{2.4}$$

Remark. When the p's are equal an expression for $P(Y_n = y)$ can be obtained from which (2.3) can be derived by elementary methods, see e.g. Feller (1968). In this case (2.1) and (2.2) are replaced by

$$N \cdot \exp(-n/N) \to m < \infty \tag{2.5}$$

or

$$n/N - \log N \to -\log m > -\infty . \tag{2.6}$$

For T_b, the number of balls until b empty cells remain, the limit distribution is given by:

Theorem 2. If b is a fixed integer and for some fixed numbers C and D,

$$0 < C \leq Np_k \leq D < \infty , \quad \text{for all } k \text{ and } N , \tag{2.7}$$

then, when $N \to \infty$,

$$\sum_{k=1}^{N} (1 - p_k)^{T_b} \Rightarrow \tfrac{1}{2} \chi^2(2(b+1)) , \tag{2.8}$$

and

$$\sum_{k=1}^{N} \exp(-T_b p_k) \Rightarrow \tfrac{1}{2} \chi^2(2(b+1)) . \tag{2.9}$$

Before proving the theorem the following functions are considered:

$$f(t) = f_N(t) = \sum_{k=1}^{N} (1-p_k)^t \ , \quad t > 0 , \tag{2.10}$$

and

(2.11) $$g(t) = g_N(t) = \sum_{k=1}^{N} \exp(-tp_k).$$

Lemma 1. If Condition (2.7) is satisfied, $y > 0$ is a fixed number, and $t = t_N = t(y)$ is defined by the equation

(2.12) $$f(t) = y ,$$

then

(2.13) $$0 < C \leq \liminf_{N \to \infty} N \log N/t_N \leq \limsup_{N \to \infty} N\log N/t_N \leq D < \infty$$

and when $N \to \infty$

(2.14) $$f([t]) \to y ,$$

(2.15) $$\max_{1 \leq k \leq N} (1 - p_k)^{[t]} \to 0 ,$$

(2.16) $$g(t) \text{ and } g([t]) \to y .$$

where $[t]$ denotes the integer part of t.

Lemma 2. If f is replaced by g and g by f in Lemma 1, then the same conclusions hold.

Proof of Lemma 1. From Condition (2.7), it follows that

(2.17) $$y = \sum_{k=1}^{N} (1 - p_k)^t \geq N \cdot (1 - D/N)^t .$$

Hence for $\varepsilon > 0$ and N sufficiently large

(2.18) $$\log y \geq \log N - t \cdot (D+\varepsilon)/N$$

and therefore

(2.19) $$D+\varepsilon = (D+\varepsilon) \lim_{N \to \infty} (1/(1-\log y/\log N)) \geq \limsup_{N \to \infty} N \log N/t_N ,$$

which proves the right inequality of (2.13).

To prove the left inequality of (2.13) the following estimate follows from (2.7):

(2.20) $$y = \sum_{1}^{N} (1-p_k)^t \leq N \cdot (1 - C/N)^t ,$$

or

$$\text{(2.21)} \qquad \log y \leq \log N - t \log(1 - C/N) \leq \log N - tC/N \,.$$

From this it follows that

$$\text{(2.22)} \qquad C = C \lim_{N\to\infty} (1 - \log y/\log N)^{-1} \leq \liminf_{N\to\infty} \, N \log N/t_N \,.$$

To prove (2.14) we observe that

$$\text{(2.23)} \qquad (1 - p_k)^{t-1} \geq (1 - p_k)^{[t]} \geq (1 - p_k)^{t} \,,$$

and using (2.7)

$$\text{(2.24)} \qquad (1 - D/N)^{-1} \sum_{1}^{N} (1-p_k)^{t} \geq \sum_{1}^{N} (1-p_k)^{[t]} \geq \sum_{1}^{N} (1 - p_k)^{t} \,,$$

or from (2.12)

$$\text{(2.25)} \qquad (1 - D/N)^{-1} y \geq f([t]) \geq y \,.$$

From which (2.14) follows.

Combining (2.7) and (2.13) give for some $K_1 > 0$ and N sufficiently large that

$$\text{(2.26)} \qquad \max(1 - p_k)^{[t]} \leq (1 - C/N)^{[t]} \leq (1 - C/N)^{K_1 N \log N} \to 0, \; N \to \infty$$

which proves (2.15).

Using (2.7) and (2.13) it follows that for some constant K

$$\text{(2.27)} \qquad |1 - e^{-tp_k}/(1 - p_k)^{t}| \leq K \cdot \log N/N \,,$$

and therefore

$$\text{(2.28)} \qquad |f(t) - g(t)| \leq \sum_{1}^{N} (1 - p_k)^{t} \cdot |1 - e^{-tp_k}/(1-p_k)^{t}|$$

$$\leq K \sum_{1}^{N} (1-p_k)^{t} \log N/N = K y \log N/N \to 0 \,,$$

which proves (2.16). ∎

Proof of Lemma 2. The proof is essentially the same as that for Lemma 1. ∎

Proof of Theorem 2. From the definitions it follows that

(2.29) $$Y_n \le b \iff T_b \le n ,$$

and therefore

(2.30) $$P(Y_n \le b) = P(T_b \le n) = P(f(T_b) \ge f(n)) .$$

Let $y > 0$ be fixed and define $n = [t]$ with $t = t(y)$ as in Lemma 1. According to Lemma 1 the assumptions of Theorem 1 are satisfied. Hence

(2.31) $$P(f(T_b) \ge y) = P(Y_n \le b) \to P(Y \le b) ,$$

where Y is $Po(y)$. Furthermore it is well-known that

(2.32) $$P(Y \le b) = P(\tfrac{1}{2}\chi^2(2(b+1)) \ge y) .$$

(2.31) and (2.32) prove (2.8). Using Lemma 2, the assertion (2.9) follows. ∎

Remark. When the p's are equal the theorem can be written

(2.33) $$N \cdot (1 - 1/N)^{T_b} \Rightarrow \tfrac{1}{2}\chi^2(2(b+1)) ,$$

and therefore

(2.34) $$T_b/N - \log N \Rightarrow \log(\tfrac{1}{2}\chi^2(2(b+1))) .$$

This result was found by Baum and Billingsley (1965) using complicated calculations. Using the result in Feller (1968) and the method of proof of Theorem 2, (2.33) and (2.34) follows. A consequence of (2.34) is

(2.35) $$T_b/N \log N \to 1 , \text{ in probability, as } N \to \infty .$$

Now (2.35) will be generalized. First introduce the distribution function

(2.36) $$H_N(x) = \#(p_k ; Np_k \le x)/N .$$

Lemma 3. If $t = t_N = t(y)$ is defined by

$$g(t) = g_N(t_N) = y > 0, \tag{2.37}$$

and there exists a distribution function $H(x)$ on $[C, D]$ such that

$$H_N(x) \to H(x), \quad N \to \infty, \tag{2.38}$$

and

$$0 < C = \inf \{x ; H(x) > 0\}, \tag{2.39}$$

then for $1/C > \varepsilon > 0$, when $N \to \infty$,

$$g_N((\varepsilon + 1/C)(N \log N)) \to 0, \tag{2.40}$$

and

$$g_N((-\varepsilon + 1/C)(N \log N)) \to +\infty. \tag{2.41}$$

Proof. From the definitions it follows that

$$\begin{aligned} 0 < y = g_N(t_N) &= N \cdot \int_C^D \exp(-t_N x/N) dH_N(x) = \\ &= \int_C^D \exp((1 - t_N x/N \log N) \log N) \, dH_N(x). \end{aligned} \tag{2.42}$$

Consider

$$g_N((\varepsilon + 1/C) N \log N) = \int_C^D \exp((1 - x(1+\varepsilon C)/C) \log N) dH_N(x). \tag{2.43}$$

Now for $C \leq x \leq D$ it is true that $1 - x(1+\varepsilon C)/C < 0$ and therefore the exponent in (2.43) is negative so the integral tend to 0 when $N \to \infty$, which proves (2.40).

For proving (2.41) consider

$$g_N((-\varepsilon + 1/C) N \log N) = \int_C^D \exp((1 - x(1 - \varepsilon C)/C) \log N) dH_N(x). \tag{2.44}$$

For $C \leq x \leq C/(1-C\varepsilon)$ the exponent is positive and as the integrand is positive (2.44) could be estimated by

$$\int_C^{C/(1-C\varepsilon)} \exp((1 - x(1-\varepsilon C)/C) \log N) dH_N(x) \to +\infty \tag{2.45}$$

by Condition (2.39). ∎

Corollary to Theorem 2. If the Conditions (2.38) and (2.39) are satisfied then

(2.46) $$T_b/N \log N \to 1/C \ , \text{ in probability, } N \to \infty \ .$$

Proof. Let $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$ be given. Take a $\delta > 0$ so that

(2.47) $$P(\tfrac{1}{2}\chi^2(2(b+1)) < \delta) < \varepsilon_2/2 \ .$$

For N sufficiently large it follows from Theorem 2 that

(2.48) $$P(g_N(T_b) < \delta) < \varepsilon_2/2$$

and from Lemma 3 that

(2.49) $$g_N((\varepsilon_1 + 1/C)(N \log N)) < \delta \ .$$

Hence

(2.50) $$P(T_b/N \log N > \varepsilon_1 + 1/C) =$$

$$P(g_N(T_b) < g_N((\varepsilon_1 + 1/C)(N \log N)) \le$$

$$P(g_N(T_b) < \delta) < \varepsilon_2/2 \ .$$

In a similar way it is proven that

(2.51) $$P(T_b/N \log N < -\varepsilon_1 + 1/C) < \varepsilon_2/2 \ .$$

Hence for N sufficiently large

(2.52) $$P(|T_b/N \log N - 1/C| > \varepsilon_1) < \varepsilon_2 \ .$$

Thus the assertion is proved. ∎

3. A small fraction of empty cells.

As above, Y_n denotes the number of empty cells after n throws.

Theorem 3. If

(3.1) $$0 < C \le Np_k \le D < \infty, \text{ for all } k \text{ and } N \ ,$$

(3.2) $$n/N \to \infty \ ,$$

and

$$(3.3)\qquad f(n) = E(Y_n) = \sum_{k=1}^{N} (1 - p_k)^n \to +\infty ,$$

then, when $n \to \infty$,

$$(3.4)\qquad (Y_n - f(n))/(f(n))^{\frac{1}{2}} \Rightarrow N(0,1) ,$$

and

$$(3.5)\qquad (Y_n - g(n))/(g(n))^{\frac{1}{2}} \Rightarrow N(0,1) ,$$

where

$$(3.6)\qquad g(n) = \sum_{k=1}^{N} \exp(-np_k) .$$

<u>Proof.</u> Using (3.1) and (3.3) it follows that

$$(3.7)\qquad \sum_{1}^{N}(1 - p_k)^n \leq N \cdot (1 - C/N)^n \to +\infty ,$$

hence

$$(3.8)\qquad n/N \log N = O(1) .$$

Using (3.1), (3.2), and (3.8) give

$$(3.9)\qquad |f(n) - g(n)| \leq \sum_{1}^{N} \exp(-np_k) \cdot$$

$$\cdot \ |\exp(n \log (1-p_k) + np_k) - 1 | \leq$$

$$\leq \sum_{1}^{N} \exp(-np_k) \cdot K \cdot n/N^2 \leq$$

$$\leq K \cdot (n/N) \cdot \exp(-C\, n/N) \to 0 .$$

Hence it is sufficient to prove (3.5). This will be established using convergence of characteristic functions.

In Holst (1971) p. 1672 the characteristic function of Y_n is given by

$$(3.10)\qquad E(\exp(i\, t\, Y_n)) = (n\,! \,/\, 2\pi i\, N^n) \cdot$$

$$\cdot \oint_{|z| = n/N} (e^{Nz}/z^{n+1}) \prod_{k=1}^{N} (1 + (e^{it}-1)\exp(-Np_k z))dz .$$

Using Stirling's formula and changing to polar coordinates it follows that

(3.11) $$E(\exp(it(Y_n - \mu)/\sigma)) = (1 + o(1)) \cdot$$

$$\cdot \int_{-\pi}^{\pi} (n/2\pi)^{\frac{1}{2}} \cdot \exp(n(e^{i\theta} - 1 - i\theta)) \cdot$$

$$\cdot \prod_{1}^{N} (\exp(-it\, e^{-np_k}/\sigma) \cdot (1 + (e^{it/\sigma} - 1)\exp(-np_k\, e^{i\theta})))d\theta$$

$$= (1 + o(1)) \cdot \int_{-\pi}^{\pi} h_n(\theta, t)d\theta ,$$

where

(3.12) $$\mu = \sigma^2 = g(n) = \sum_{1}^{N} \exp(-np_k), \quad \sigma > 0 .$$

The integral will be studied by the same method as in Holst (1971).

Take $0 < a < 1/6$ and split the interval $-\pi \leq \theta \leq \pi$ into

(3.13) $$A = \{\theta \, ; \, a \leq |\theta| \leq \pi \} ,$$

(3.14) $$B = \{ \theta ; \, n^{a-\frac{1}{2}} \leq |\theta| < a \} ,$$

and

(3.15) $$C = \{\theta \, ; |\theta| < n^{a-\frac{1}{2}} \} .$$

From Lemmas 4-6 below it follows that

(3.16) $$E(\exp(it(Y_n - \mu)/\sigma) = (1 + o(1)) \cdot$$

$$(\int_A h_n + \int_B h_n + \int_C h_n) \to 0 + 0 + \exp(-t^2/2), \quad n \to \infty .$$

By the continuity theorem for characteristic functions assertion (3.5) is proved, and thus the theorem. ∎

With the same conditions as in Theorem 3 the following lemmas hold.

<u>Lemma 4.</u> For every fixed real number t

(3.17) $$\int_A h_n(\theta,t)d\theta \to 0 \ , \quad n \to \infty \ .$$

<u>Proof.</u> As $n/N \to \infty$ and $\sigma \to \infty$ it follows that

(3.18) $$\left|\int_A\right| \leq K_1 \cdot n^{\frac{1}{2}} e^{-n} \cdot \int_A \prod_1^N |\exp(np_k e^{i\theta}) + e^{it/\sigma} - 1| d\theta$$

$$\leq K_2 n^{\frac{1}{2}} e^{-n} \prod_1^N (\exp(np_k \cos a) + o(1))$$

$$\leq K_2 n^{\frac{1}{2}} e^{-n} 2^N e^{n\cos a} \to 0 \ . \qquad \blacksquare$$

<u>Lemma 5</u> For every fixed real number t

(3.19) $$\int_B h_n(\theta,t)d\theta \to 0 \ , \quad n \to \infty \ .$$

<u>Proof.</u> From the assumptions, it follows that there exist positive numbers $K_3 - K_9$ such that

(3.20) $$\left|\int_B\right| \leq K_3 n^{\frac{1}{2}} e^{-n} \int_B \prod_1^N (\exp(np_k \cos\theta) + O(1/\sigma))d\theta$$

$$\leq K_4 n^{\frac{1}{2}} e^{-n} \prod_1^N \exp(np_k \cos n^{a-\frac{1}{2}}) \cdot$$

$$\cdot (1 + K_5 \cdot \exp(-K_6 n/N)/\sigma)$$

$$\leq K_7 n^{\frac{1}{2}} e^{-n} \exp(n(1 - K_8 n^{2a-1}))$$

$$\leq \exp(-K_9 n^{2a}) \to 0 \ , \quad n \to \infty \ . \qquad \blacksquare$$

<u>Lemma 6.</u> For every fixed real number t,

(3.21) $$\int_C h_n(\theta,t)d\theta \to \exp(-t^2/2), \quad n \to \infty \ .$$

<u>Proof.</u> Expanding in series gives

(3.22) $$\log h_n(\theta,t) = -n\,\theta^2/2 + o(1)$$

$$+ \sum_1^N (\log(1 + \exp(-np_k e^{i\theta})(e^{it/\sigma} - 1)) - it\exp(-np_k)/\sigma) + \tfrac{1}{2}\log(n/2\pi).$$

Now, when $n \to \infty$,

$$(3.23) \qquad \sum_1^N \left| \exp(-2np_k\, e^{i\theta})(e^{it/\sigma} - 1)^2 \right|$$

$$= o(1) \cdot \sum_1^N \exp(-np_k)/\sigma^2 = o(1) ,$$

and therefore

$$(3.24) \qquad \sum_1^N (\log (1 + \ldots) - \ldots)$$

$$= \sum_1^N (\exp(-np_k\, e^{i\theta})(e^{it/\sigma} - 1) - it \exp(-np_k)/\sigma) + o(1) .$$

Furthermore, using (3.8), (3.9) and the assumptions, it follows that

$$(3.25) \qquad \sum_1^N \exp(-np_k\, e^{i\theta})/\sigma^2 \to 1 ,$$

and therefore (3.24) can be written

$$(3.26) \qquad \sum_1^N (\ldots) = \sum_1^N (\exp(-np_k\, e^{i\theta})(it/\sigma - t^2/2\sigma^2)$$

$$- it \exp(-np_k)/\sigma) + o(1)$$

$$= it \sum_1^N (\exp(-np_k(e^{i\theta} - 1)) - 1) \exp(-n p_k)/\sigma$$

$$- t^2/2 + o(1) .$$

Now, when $n \to \infty$,

$$(3.27) \qquad \sum_1^N (np_k)^2\, \theta^2 \exp(-np_k)/\sigma \leq$$

$$\leq K_1 (n/N)^2\, n^{2a-1}\, N^{\frac{1}{2}} \exp(-K_2\, n/N) \to 0 .$$

From this it follows that

$$(3.28) \qquad \sum_1^N (\ldots) = \theta t \sum_1^N n p_k \exp(-np_k)/\sigma - t^2/2 + o(1) .$$

Hence for θ in C,

(3.29) $$\log h_n(\theta,t) - \tfrac{1}{2}\log(2\pi/n) = -n\theta^2/2 + \theta t \sum_1^N np_k \exp(-np_k)/\sigma$$
$$- t^2/2 + o(1) = -(n^{\frac{1}{2}}\theta - t\sum_1^N n^{\frac{1}{2}} p_k \exp(-np_k)/\sigma)^2/2$$
$$-t^2(1-(\sum_1^N n^{\frac{1}{2}} p_k \exp(-np_k)/\sigma)^2)/2 + o(1).$$

Now, when $n \to \infty$,

(3.30) $$\sum_1^N n^{\frac{1}{2}} p_k \exp(-np_k)/\sigma \le K_3 n^{\frac{1}{2}} N^{-1} \cdot N^{\frac{1}{2}}$$
$$\cdot \exp(-K_4 n/N) \to 0.$$

Thus with $\psi = n^{\frac{1}{2}}\theta$ the integral (3.21) can be written

(3.31) $$\int_C h_n = \int_{|\psi| \le n^a} (2\pi)^{-\frac{1}{2}}$$
$$\cdot \exp(-(\psi - o(1))^2/2 - t^2/2 + o(1))\, d\psi,$$

which converges to $\exp(-t^2/2)$ when $n \to \infty$. ∎

4. The waiting time for a small fraction.

As above let T_b denote the number of balls thrown until exactly $b = b_N$ cells remain empty. Let t_b be the unique solution of the equation

(4.1) $$b = g(t_b) = \sum_{k=1}^N \exp(-t_b p_k).$$

Theorem 4. If, when $N \to \infty$,

(4.2) $$b_N \to +\infty,$$

(4.3) $$b_N/N \to 0,$$

and

(4.4) $$0 < C \le Np_k \le D < \infty, \text{ for all } k \text{ and } N,$$

then

(4.5) $$b_N^{-\frac{1}{2}}(T_b - t_b)\sum_{k=1}^N p_k \exp(-t_b p_k) \Rightarrow N(0,1).$$

Proof. From the assumptions it follows that

$$C\, b/N \le \Delta = \sum_{1}^{N} p_k \exp(-t_b p_k) \le D\, b/N . \tag{4.6}$$

Thus for N sufficiently large

$$0 < C \le \Delta \cdot N/b \le D < \infty . \tag{4.7 }$$

As in the proof of Theorem 2 the following relation holds

$$P((T_b - t_b)\, \Delta/b^{\frac{1}{2}} \le x) = P(Y_n \le b) , \tag{4.8}$$

where

$$n = [t_b + x\, b^{\frac{1}{2}}/\Delta] . \tag{4.9}$$

It is seen that

$$\begin{aligned} g(n)\,(1 + o(1)) &= g(t_b + x\, b^{\frac{1}{2}}/\Delta) \\ &= \sum \exp(-t_b p_k) \cdot (1 - x\, p_k\, b^{\frac{1}{2}}/\Delta + O(1/b)) \\ &= b - x \cdot b^{\frac{1}{2}} + O(1) , \end{aligned} \tag{4.10}$$

and thus

$$g(n) \to +\infty , \tag{4.11}$$

and from (3.9) it follows that

$$f(n) \to +\infty . \tag{4.12}$$

Furthermore,

$$b = g(t_b) \ge N \exp(-D\, t_b/N), \tag{4.13}$$

implying that

$$t_b/N \to +\infty , \tag{4.14}$$

and therefore

$$n/N \to +\infty . \tag{4.15}$$

Hence the assumptions of Theorem 3 are fulfilled and (4.8) and (4.10) give

$$P((T_b - t_b)/b^{\frac{1}{2}} \le x) = P(Y_n \le b) = \tag{4.16}$$

$$= \Phi((b - g(n)) / (g(n))^{\frac{1}{2}}) + o(1) =$$

$$= \Phi((x b^{\frac{1}{2}} + O(1))/(b(1 + o(1)))^{\frac{1}{2}}) + o(1) \to \Phi(x) ,$$

where $\Phi(x)$ is the standardized normal distribution function. This proves the theorem. ∎

References

Baum, L. E. and Billingsley, P. (1965), Asymptotic distributions for the coupon collector's problem. Ann. Math. Statist. 36, 1835-1839.

Feller, W. (1968), An Introduction to Probability Theory and its Applications, 1, 3rd ed., Wiley, New York.

Holst, L., (1971), Limit theorems for some occupancy and sequential occupancy problems. Ann. Math. Statist. 42, 1671-1680.

Samuel-Cahn, E. (1974), Asymptotic distributions for occupancy and waiting time problems with positive probability of falling through the cells. Ann. Probability 2, 515-521.

Sevastyanov, B. A. (1972), Poisson limit law for a scheme of sums of dependent random variables, Theor. Probability Appl. 17, 695-699.